Machine Learning 1

Lecture 3.4 - Supervised Learning
Underfitting and Overfitting

Erik Bekkers

(Bishop 1.1)

Image credit: Kirillm | Getty Images

Slide credits: Rianne van den Berg

Example: Overfitting and Underfitting

$t = \sin(2\pi x) + \varepsilon \qquad \varepsilon \sim \mathcal{N}(0, \beta^{-1}) \qquad y(x, \mathbf{w}) = w_0 + \sum_{i=1}^{M} w_i x^i$

Underfitting

overfitting

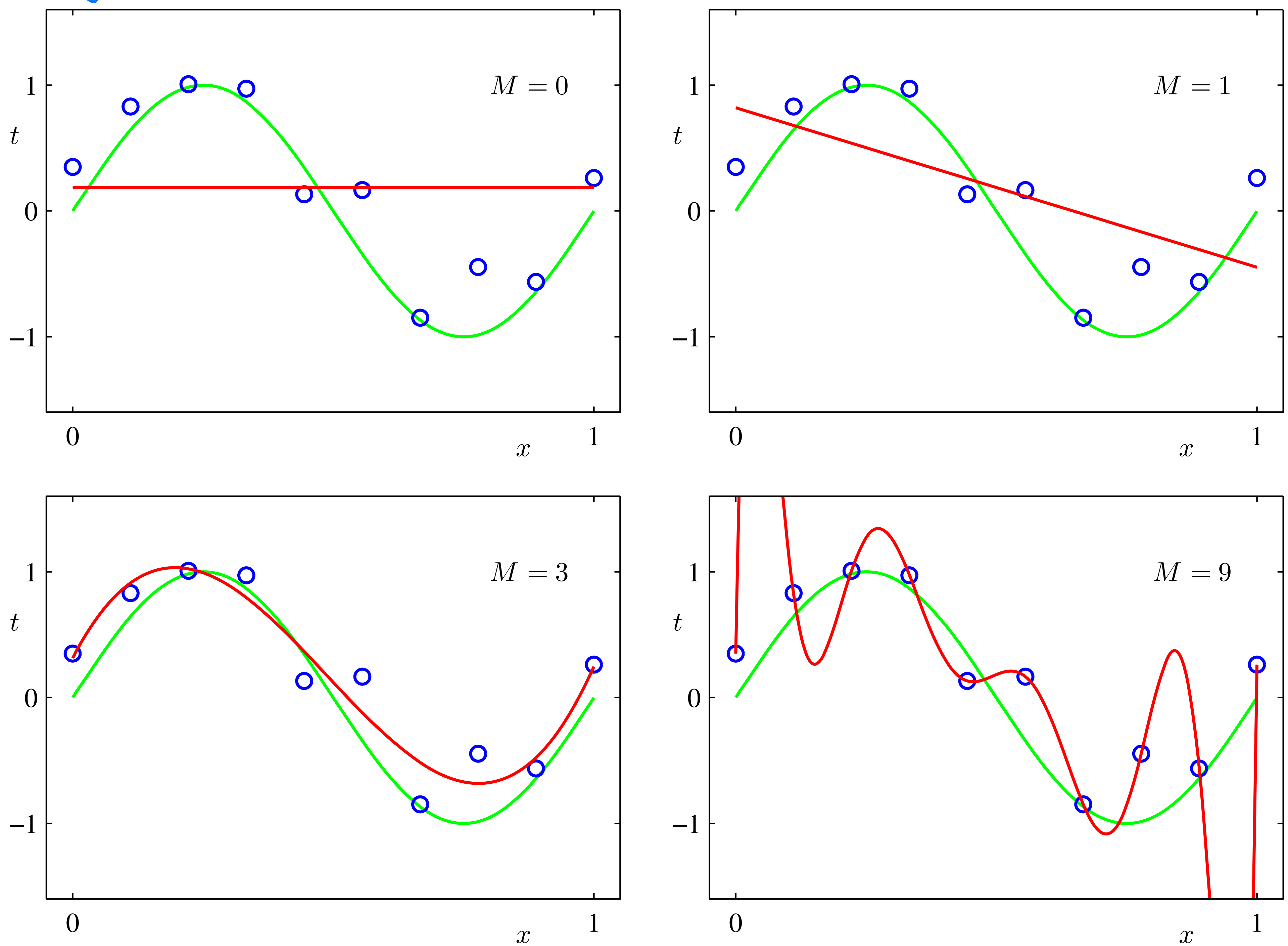


Figure: Fits of different polynomials (Bishop 1.4)

Example: Overfitting and Underfitting

M=6 spans a subspace of M=9

	$M = 0$	$M = 1$	$M = 6$	$M = 9$
$w_0^\star$	0.19	0.82	0.31	0.35
$w_1^\star$		-1.27	7.99	232.37
$w_2^\star$			-25.43	-5321.83
$w_3^\star$			17.37	48568.31
$w_4^\star$				-231639.30
$w_5^\star$				640042.26
$w_6^\star$				-1061800.52
$w_7^\star$				1042400.18
$w_8^\star$				-557682.99
$w_9^\star$				125201.43

So why not set this part to zero?

Sign of overfitting

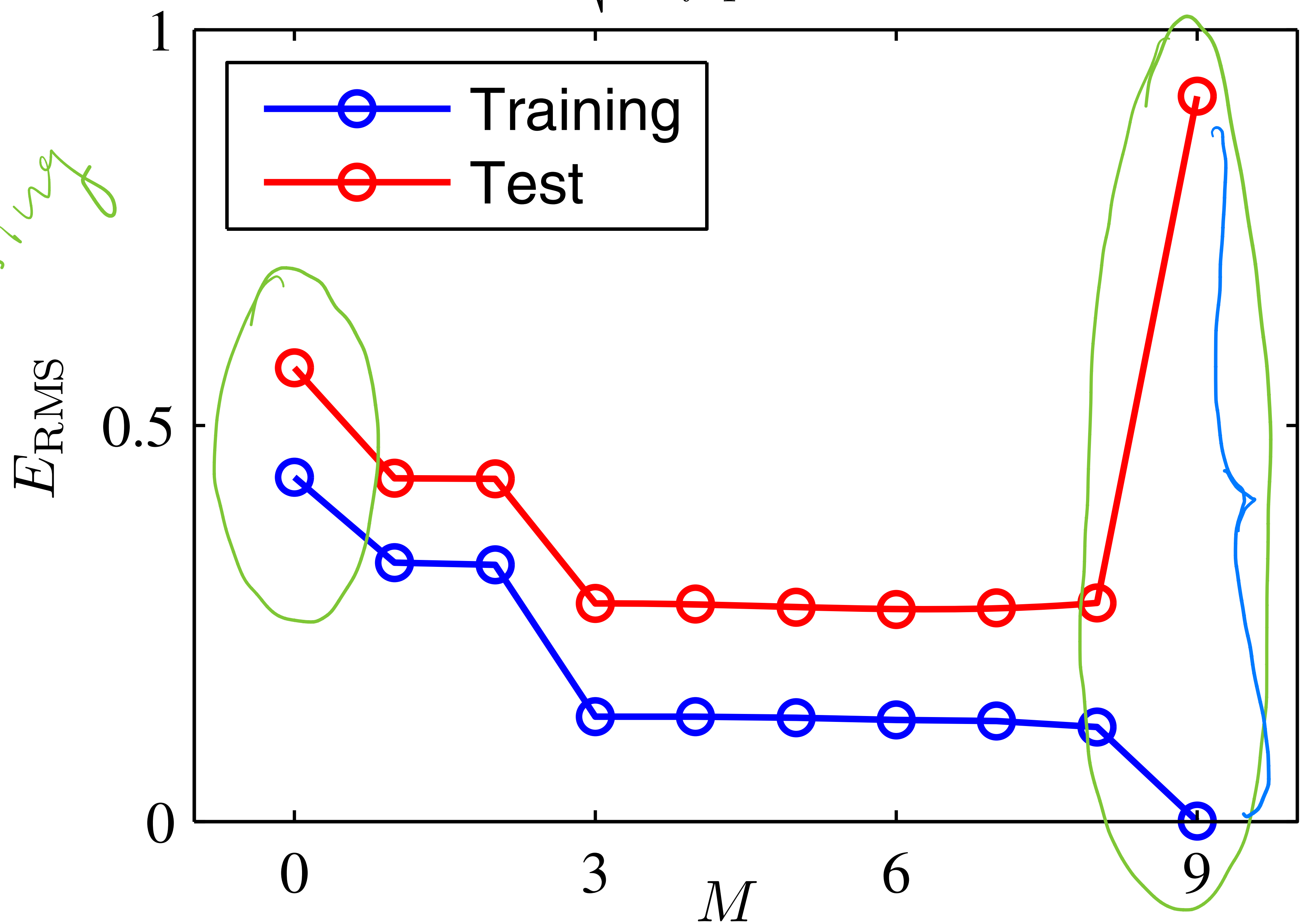
Because or model is asked to minimize the sum of squared errors!

Table: Polynomial coefficients (Bishop 1.1)

Example: Overfitting and Underfitting

$$E_{\mathrm{RMS}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\{t_i - y(x, \mathbf{w})\}^2}$$

underfitting



overfitting

generalization gap

Figure: training and test RMSE (Bishop 1.5)

Example: Overfitting

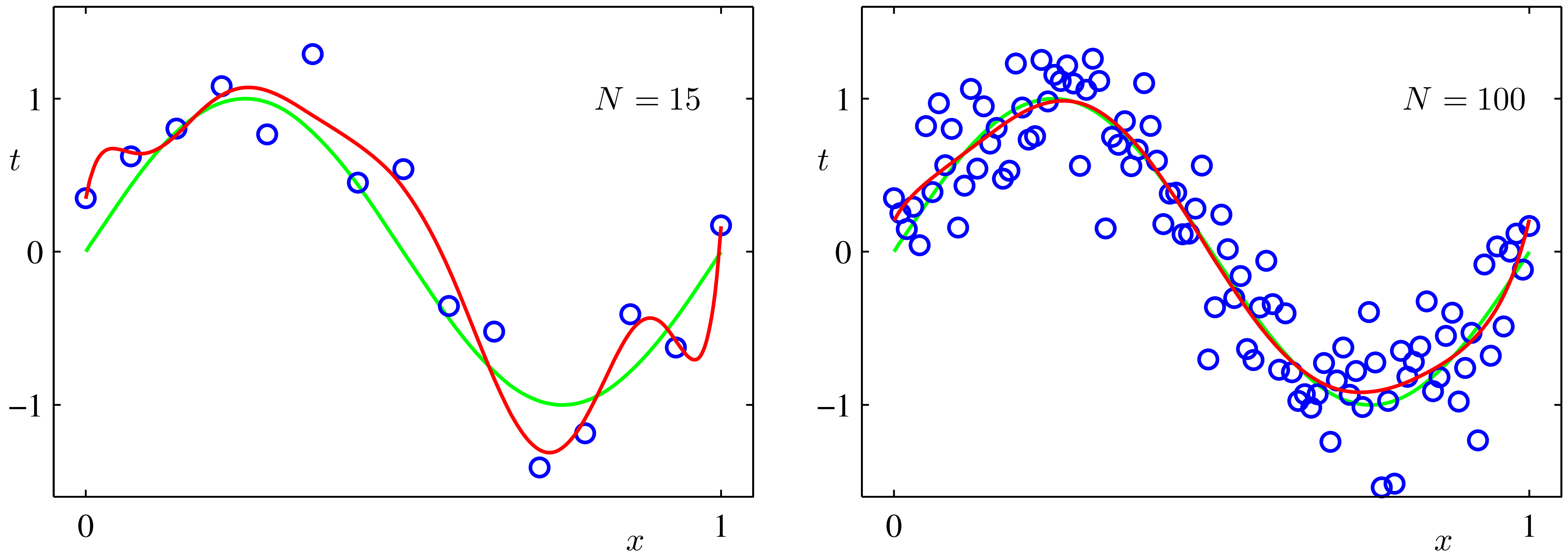


Figure: *M*=9 Polynomial fit with increased datapoints N. (Bishop 1.6)